

# Stochastic multireference Epstein-Nesbet perturbation theory

Sandeep Sharma,<sup>1,2,\*</sup> Adam Holmes,<sup>3,†</sup> Guillaume Jeanmairet,<sup>1</sup> Ali Alavi,<sup>1,4,‡</sup> and C. J. Umrigar<sup>3,§</sup>

<sup>1</sup>*Max Planck Institute for Solid State Research, Heisenbergstraße 1, 70569 Stuttgart, Germany*

<sup>2</sup>*Department of Chemistry and Biochemistry, University of Colorado Boulder, Boulder, CO 80302, USA*

<sup>3</sup>*Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY 14853, USA*

<sup>4</sup>*Dept of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom*

We extend the recently proposed heat-bath configuration interaction (HCI) method [Holmes, Tubman, Umrigar, *J. Chem. Theory Comput.* **12**, 3674 (2016)], by introducing a stochastic algorithm for performing multireference Epstein-Nesbet perturbation theory, in order to completely eliminate the severe memory bottleneck of the original method. The proposed stochastic algorithm has several attractive features. First, there is no sign problem that plagues several quantum Monte Carlo methods. Second, instead of using Metropolis-Hastings sampling, we use the Alias method to directly sample determinants from the reference wavefunction, thus avoiding correlations between consecutive samples. Third, in addition to removing the memory bottleneck, stochastic-HCI (s-HCI) is faster than the deterministic variant for most systems if a stochastic error of 0.1 mHa is acceptable. Fourth, within the s-HCI algorithm one can trade memory for a modest increase in computer time. Fifth, the perturbative calculation is embarrassingly parallel. The s-HCI algorithm extends the range of applicability of the original algorithm, allowing us to calculate the correlation energy of very large active spaces. We demonstrate this by performing calculations on several first row dimers including F<sub>2</sub> with an active space of (14e, 108o), Mn-Salen cluster with an active space of (28e, 220o), and Cr<sub>2</sub> dimer with up to a quadruple-zeta basis set with an active space of (12e, 190o). For these systems we were able to obtain better than 1 mHa accuracy with a wall time of merely 113 seconds, 65 seconds, and 3 hours on 1, 1, and 8 nodes, respectively.

## I. INTRODUCTION

Many methods, e.g., coupled cluster and Møller-Plesset perturbation theory, can accurately and efficiently treat the electronic correlation of single-reference (weakly-correlated) systems. In particular, coupled cluster with singles, doubles and perturbative triples (CCSD(T)) is very accurate for such systems and is often referred to as the “gold standard” of quantum chemistry. However, these methods fail catastrophically when applied to multireference (strongly-correlated) systems, such as molecules undergoing chemical reactions or systems containing transition metal atoms with partially filled *d* or *f* orbitals.

One common approach for tackling such multireference problems is to abandon the Hartree-Fock wavefunction and instead use a multideterminantal reference wavefunction obtained by correlating a subset of orbitals around the Fermi surface. Examples include the complete active space (CAS) method, in which all possible occupancies of orbitals within the active space are included, and the restricted/generalized active space (RAS/GAS) methods<sup>1–3</sup>, in which further restrictions are placed on the occupancies of the active orbitals in order to reduce the size of the Hilbert space. The CAS method is limited to about 16 active electrons and orbitals. Other possibilities include highly accurate but approximate methods such as the density matrix renormalization group (DMRG)<sup>4–6</sup>, full configuration interaction quantum Monte Carlo (FCIQMC)<sup>7,8</sup>, and its semistochastic improvement (S-FCIQMC)<sup>9</sup>, which routinely treat up to about 40-50 active orbitals. A well-chosen active space often results in a reference wavefunction that contains

qualitatively correct physics. However, quantitative accuracy requires one to take into account the dynamical correlation by allowing excitations into inactive-space orbitals. Common methods for including dynamical correlation include multireference configuration interaction (MRCI) and its size-consistent variants<sup>10–12</sup>, various flavors of multireference perturbation theory<sup>13–16</sup>, and multireference coupled cluster theories<sup>17–20</sup>. The accuracy of these methods is often limited by the fact that only a relatively small number of active space orbitals can be used to obtain the reference wavefunction because the cost of enlarging the active space increases exponentially with the number of orbitals.

Although the number of determinants in a CAS scales combinatorially with the number of active electrons and orbitals, many of these determinants are “configurational deadwood,” and do not contribute appreciably to the reference wavefunction. The so-called *selected configuration interaction* (SCI) methods, which have been in use for more than four decades<sup>21–35</sup>, take advantage of this fact and generate a reference wavefunction by selecting only important determinants, rather than including all determinants in the CAS. A subset of these methods improve upon the variational energy by employing a perturbative correction to the energy using multireference Epstein-Nesbet perturbation theory. We refer to these methods as *selected configuration interaction plus perturbation theory* (SCI+PT) methods. The first such method was called *configuration interaction perturbing a multi-configurational zeroth-order wavefunction selected iteratively* (CIPSI)<sup>21</sup>.

The focus of this paper is on a newly-introduced SCI+PT method called *heat-bath configuration interac-*

tion (HCI). HCI distinguishes itself from other SCI+PT techniques by employing an algorithm that greatly improves the efficiency of both the variational and perturbative steps. Although it is more efficient than other SCI+PT methods, HCI, in its original formulation, is limited by a memory bottleneck because it stores in memory all the determinants that contribute to the perturbative correction<sup>36</sup>.

In this paper, we introduce a stochastic implementation of multireference Epstein-Nesbet perturbation theory, and use it to overcome the memory bottleneck of HCI. This method has several attractive properties. First, it does not have a sign problem that plagues quantum Monte-Carlo methods. Second, instead of using the Metropolis-Hastings method, we use the Alias method, so the samples are all uncorrelated. Third, in addition to removing the memory bottleneck, stochastic-HCI (s-HCI) is faster than the deterministic variant for even the smallest system tried in the current work if a stochastic error of 0.1 mHa is acceptable. Fourth, within the s-HCI algorithm one can trade memory for a modest increase in computer time. Fifth, the perturbative calculation is embarrassingly parallel.

In Section II we review the improvements made in the original HCI algorithm that make it much more efficient than other SCI+PT algorithms. In Section III, we present our stochastic perturbation theory which removes the memory bottleneck of the original HCI algorithm. In Section IV we provide various implementation details of both the variational and the perturbative parts of our algorithm. We then demonstrate the utility of the stochastic-HCI (s-HCI) method by applying it in Section V to various diatomic molecules including  $F_2$  with an active space of (14e, 108o), Mn-Salen cluster with an active space of (28e, 22o), and  $Cr_2$  dimer with up to a quadruple-zeta basis set with an active space of (12e, 190o), obtaining energies that are accurate to better than 1 mHa with very modest computer resources. Finally, in Section VI, we conclude and discuss future research directions.

## II. HEAT-BATH CONFIGURATION INTERACTION

We begin by describing the HCI algorithm in its original formulation<sup>37</sup>, emphasizing the key innovations that make it much more efficient than other SCI+PT methods. In the following discussion the indices  $i, j, \dots$  will be used for determinants in the variational space  $\mathcal{V}$  and the indices  $a, b, \dots$  will be used for determinants in  $\mathcal{C}$ , the space of determinants that are connected to  $\mathcal{V}$  but not in  $\mathcal{V}$ . Similar to other SCI+PT methods, HCI has two stages: (1) a variational stage, in which a variational wavefunction is obtained as a linear combination of a set of determinants chosen by an iterative procedure, and (2) a perturbative stage, in which the second-order correction to the variational energy is computed using mul-

tireference Epstein-Nesbet perturbation theory<sup>38,39</sup>, but each stage is much faster than in other SCI+PT methods.

### Variational Stage

At the start of the algorithm,  $\mathcal{V}$  consists of some initial set of determinants, usually just the Hartree-Fock determinant. Then, at each iteration, new determinants are added to  $\mathcal{V}$ , chosen using a parameter  $\epsilon_1$ , as follows. The initial wavefunction is the ground state of the Hamiltonian in  $\mathcal{V}$ ,  $|\Psi_0\rangle = \sum_i c_i |D_i\rangle$ . At each iteration:

1. Add to the variational space  $\mathcal{V}$ , all determinants  $D_a$  in the space of connections  $\mathcal{C}$ , such that

$$|H_{ai}c_i| > \epsilon_1 \quad (1)$$

for at least one determinant  $D_i$  in the current  $\mathcal{V}$ .

2. Calculate the lowest eigenvalue  $E_0$  with eigenvector  $|\Psi_0\rangle = \sum_i c_i |D_i\rangle$  of the Hamiltonian in  $\mathcal{V}$ .

The iterations are terminated when the number of new determinants is less than a threshold, e.g., 1% of the current size of  $\mathcal{V}$ , or when a maximum number of iterations is reached. Since the values of  $c_i$  tend to be larger in the initial iterations when there are few determinants in  $\mathcal{V}$ ,  $\epsilon_1$  is set during the first few iterations to be larger than its final value.

HCI takes advantage of the fact that the double excitation matrix elements depend only on the four orbitals whose occupancy is changing. Step 1 is performed efficiently by storing the double excitation matrix elements in order of decreasing magnitude, so that no time is wasted on determinants that do not meet the cutoff in Eq. 1. For details, we refer the reader to the original HCI paper<sup>37</sup>. Thus, we see that HCI identifies new determinants to add to  $\mathcal{V}$  in a manner that is more efficient than other SCI methods in two ways:

- HCI uses a selection criterion which is *cheap to evaluate* for each determinant, namely Eq. 1. In contrast, other SCI methods use a criterion based on a perturbative expression which is more expensive; for example, CIPSI<sup>21</sup> uses the magnitude of the coefficient of the first-order correction to the wavefunction, namely  $\left| \frac{\sum_i H_{ai}c_i}{E_0 - H_{aa}} \right| > \epsilon_1$ .
- HCI evaluates its selection criterion (Eq. 1) *only for those determinants which will be added to  $\mathcal{V}$ !* By comparison, other SCI methods iterate through *all* candidate determinants  $\{D_a\}$  (determinants for which there exists at least one nonzero matrix element  $H_{ai}$  with  $D_i \in \mathcal{V}$ ), evaluating their expensive selection criteria for each one.

The simplification in HCI is possible because it was demonstrated<sup>37</sup> that variation in the perturbative expression for the coefficients is dominated by variation in the largest-magnitude term in the numerator, since the matrix elements  $\{H_{ai}\}$  and coefficients  $\{c_i\}$  span many orders of magnitude. The minor deviation from optimality

in the choice of the most important determinants is by far outweighed by the fact that many more determinants can be included.

#### Perturbative Stage

The variational wavefunction is used to define the zeroth order Hamiltonian,  $H_0$  and the perturbation,  $V$ ,

$$H_0 = \sum_{i,j} H_{ij} |D_i\rangle\langle D_j| + \sum_a H_{aa} |D_a\rangle\langle D_a|.$$

$$V = H - H_0 \quad (2)$$

It can easily be verified that  $|\Psi_0\rangle$  is the ground state of  $H_0$  with eigenvalue  $E_0$ . Using the partitioning in Eq. 2, the first order correction of the wavefunction  $|\Psi_1\rangle$  and the second order energy correction  $\Delta E^{(2)}$  can be written as

$$|\Psi_1\rangle = \frac{1}{E_0 - H_0} V |\Psi_0\rangle$$

$$= \sum_a \frac{\sum_i H_{ai} c_i}{E_0 - E_a} |D_a\rangle \quad (3)$$

and

$$\Delta E^{(2)} = \langle \Psi_0 | V | \Psi_1 \rangle$$

$$= \sum_a \frac{(\sum_i H_{ai} c_i)^2}{E_0 - E_a}, \quad (4)$$

where  $E_a = H_{aa}$ . It is worth noting that the expression for the total energy,  $E_0 + \Delta E^{(2)}$  is identical to that for the mixed estimator of the energy used in quantum Monte Carlo calculations, provided that the projected wavefunction is replaced by the perturbed wavefunction.

This expression in Eq. 4 is expensive to calculate, as it requires a summation over many small terms. Instead, HCI includes only those terms in the sum that contribute substantially,

$$\Delta E^{(2)} \approx \sum_a \frac{\left( \sum_i^{(\epsilon_2)} H_{ai} c_i \right)^2}{E_0 - E_a}, \quad (5)$$

where  $\sum^{(\epsilon_2)}$  denotes a sum in which all terms in the sum that are smaller in magnitude than  $\epsilon_2$  are discarded, i.e.,  $\sum_i^{(\epsilon_2)} H_{ai} c_i$  includes only terms for which  $|H_{ai} c_i| > \epsilon_2$ .

Once again, since the double excitation matrix elements are stored in order of decreasing magnitude, no time is spent on terms that do not contribute to the sum. The parameter  $\epsilon_2$  is kept much smaller than the parameter  $\epsilon_1$  because discarding small amplitude determinants can lead to significant errors in the calculation of dynamical correlation. In the original HCI paper<sup>37</sup>, for each  $\epsilon_1$ , several values of  $\epsilon_2$  were used, and the energy for  $\epsilon_2 = 0$  was obtained by extrapolation. In this paper, a single value of  $\epsilon_2$  is used, that is sufficiently small to recover the  $\epsilon_2 = 0$  limit to a precision that is better than 1 mHa.

It was shown in the previous publication<sup>37</sup> that the above algorithm is highly efficient and can be used to obtain sub-milliHartree accuracy for challenging problems like all-electron (48e,42o) Cr<sub>2</sub> with the small Ahlrichs double-zeta basis<sup>40</sup> in a few minutes on a single computer core. However, since the contributions from all  $i$  in Eq. 5 must be summed and then squared, the efficient deterministic approach to computing the perturbative correction requires storing the partial sums  $\left\{ \sum_i^{(\epsilon_2)} H_{ai} c_i \right\}$  for all  $a$  for which  $|H_{ai} c_i| > \epsilon_2$  which creates a severe memory bottleneck<sup>36</sup>. In the following section we show how by using a stochastic version of the perturbation theory this memory bottleneck can be completely eliminated.

### III. STOCHASTIC MULTIREFERENCE PERTURBATION THEORY

We write the perturbative correction in a slightly different form than presented in Eq (5) to highlight the fact that it is a bilinear function of the coefficients of the zeroth-order state.

$$\Delta E^{(2)} = \sum_a \frac{1}{E_0 - E_a} \left( \sum_{ij}^{(\epsilon_2)} H_{ai} H_{aj} c_i c_j \right). \quad (6)$$

We compute the expected value of this expression stochastically by performing  $N_b$  iterations; in each, we sample the zeroth order wavefunction by selecting  $N_d$  determinants  $\{D_i\}$  from  $\mathcal{V}$  each with probability

$$p_i = \frac{|c_i|}{\sum_i |c_i|}. \quad (7)$$

Any given sample will contain  $N_d^{\text{diff}}$  distinct determinants  $D_i$  with some number of repeats  $w_i$ , such that

$$\sum_i^{N_d^{\text{diff}}} w_i = N_d$$

The number of repetitions ( $w_i$ ) is distributed according to the well known multinomial distribution. The mean and second moment, for  $i \neq j$ , of this distribution are

$$\langle w_i \rangle = p_i N_d \quad (8)$$

$$\langle w_i w_j \rangle = p_i p_j N_d (N_d - 1), \quad (9)$$

where  $\langle \cdot \rangle$  denotes the expectation value of a quantity evaluated for a sample of  $N_d$  determinants, a notation we will use hereafter.

Using these expressions, the unbiased estimate of the second order perturbation can be calculated from the sampled wavefunction as follows,

$$\begin{aligned}
\Delta E^{(2)} &= \sum_a \frac{1}{E_0 - E_a} \left[ \sum_{ij}^{\mathcal{V}} H_{ai} H_{aj} c_i c_j \right] \\
&= \sum_a \frac{1}{E_0 - E_a} \left[ \sum_{i \neq j}^{\mathcal{V}} H_{ai} H_{aj} c_i c_j + \sum_i^{\mathcal{V}} H_{ai}^2 c_i^2 \right] \\
&= \left\langle \sum_a \frac{1}{E_0 - E_a} \left[ \sum_{i \neq j}^{N_d^{\text{diff}}} \frac{w_i w_j c_i c_j H_{ai} H_{aj}}{\langle w_i w_j \rangle} + \sum_i^{N_d^{\text{diff}}} \frac{w_i c_i^2 H_{ai}^2}{\langle w_i \rangle} \right] \right\rangle \\
&= \left\langle \sum_a \frac{1}{E_0 - E_a} \left[ \sum_{i \neq j}^{N_d^{\text{diff}}} \frac{w_i w_j c_i c_j H_{ai} H_{aj}}{p_i p_j N_d (N_d - 1)} + \sum_i^{N_d^{\text{diff}}} \frac{w_i c_i^2 H_{ai}^2}{p_i N_d} \right] \right\rangle \\
&= \frac{1}{N_d (N_d - 1)} \left\langle \sum_a \frac{1}{E_0 - E_a} \left[ \left( \sum_i^{N_d^{\text{diff}}} \frac{w_i c_i H_{ai}}{p_i} \right)^2 + \sum_i^{N_d^{\text{diff}}} \left( \frac{w_i (N_d - 1)}{p_i} - \frac{w_i^2}{p_i^2} \right) c_i^2 H_{ai}^2 \right] \right\rangle, \quad (10)
\end{aligned}$$

where for brevity we have suppressed the superscript ( $\epsilon_2$ ) on the  $i$  and  $j$  sums, though of course we will always use a nonzero  $\epsilon_2$  value for efficiency.

Going from the 2<sup>nd</sup> to the 3<sup>rd</sup> line above, we replace the sum over the states in  $\mathcal{V}$  by a sum over the sample, so in order to have an unbiased expectation value we divide the two terms by  $\langle w_i w_j \rangle$  and  $\langle w_i \rangle$  respectively. In going from the 3<sup>rd</sup> to the 4<sup>th</sup> line we use Eqs. 8 and 9.

In practice, the exact average in Eq. (10) will be replaced by an average over  $N_b$  iterations. For any  $N_d \geq 2$  we obtain an unbiased estimate of the second order correction to the energy and this estimate can be made progressively more precise by averaging over a large number of iterations  $N_b$ . Each batch contains an independently chosen set of  $N_d^{\text{diff}}$  determinants and thus there is no autocorrelation between consecutive batches. This is in sharp contrast to discrete-space quantum Monte Carlo methods, such as the FCIQMC method<sup>7,8</sup> and its semistochastic improvement<sup>9</sup>, for which the autocorrelation time increases both with system size and the size of the basis to the point that it can become difficult to accurately estimate the statistical error. This drawback of the FCIQMC method is ameliorated but not eliminated by using the more efficient sampling method of Ref. 41.

We note that the expression in Eq. (10) is evaluated in much the same way as the deterministic evaluation of the perturbative correction using a single batch, the main difference being that the  $N_v$  variational determinants have been replaced by the much smaller subset of  $N_d^{\text{diff}}$  distinct sampled determinants and that an additional summation is needed to ensure that the result is unbiased. It is important to note that for a single batch, the summation over  $a$  in Eq. (10) is restricted to only those determinants in  $\mathcal{C}$  that have a non-zero Hamiltonian matrix element with the  $N_d^{\text{diff}}$  determinants used to sample the zeroth order wavefunction.

Figure 1a shows that the CPU time per sample increases nearly linearly with  $N_d$ , the number of determinants in the sample, for the C<sub>2</sub> and F<sub>2</sub> molecules. As shown in Section IV, the scaling contains two terms one that scales linearly with  $N_d$  and another that scales as  $N_d \log(N_d)$  (Ref. 42). Figure 1b shows the CPU time necessary to reach a standard deviation of less than 0.1 mHa versus the number of determinants in the sampled wavefunction  $N_d$ . There is a rapid initial decrease followed by a much shallower decrease beyond about  $N_d = 200$ . Consequently, it is desirable to use as large a value of  $N_d$  as memory allows, but the gain from using  $N_d > 200$  is minor. Another consideration is that  $N_b$  needs to be large enough to get a reasonable estimate of the statistical error, and since the computer time is approximately  $\propto N_d N_b$ , it sometimes makes sense to use a smaller  $N_d$  than available memory allows. In all the calculations presented in Section V, we have used  $N_d = 200$ .

It is worth mentioning that the memory bottleneck can also be removed without recourse to the stochastic method. This can be achieved by dividing the  $N_v$  determinants in  $\mathcal{V}$  into  $N_b$  batches, each containing on average  $N_d$  determinants ( $N_b = N_v/N_d$ ) since all determinants in  $\mathcal{V}$  need to be in a batch, and computing the contribution from all pairs of batches independently. For large systems the  $N_b$  required to get a statistical error of 1 mHa in the stochastic method is much smaller than the  $N_b$  required in the deterministic calculation, making the stochastic approach the more efficient choice. In Section IV we will see that the leading cost of performing the calculation for each pair of batches is  $\propto N_d$  and so the cost of performing the entire calculation containing  $N_b$  batches is  $\propto N_d N_b^2 \propto N_v N_b$ . Thus the cost of the deterministic calculation scales linearly with  $N_b$  and quickly becomes very expensive.

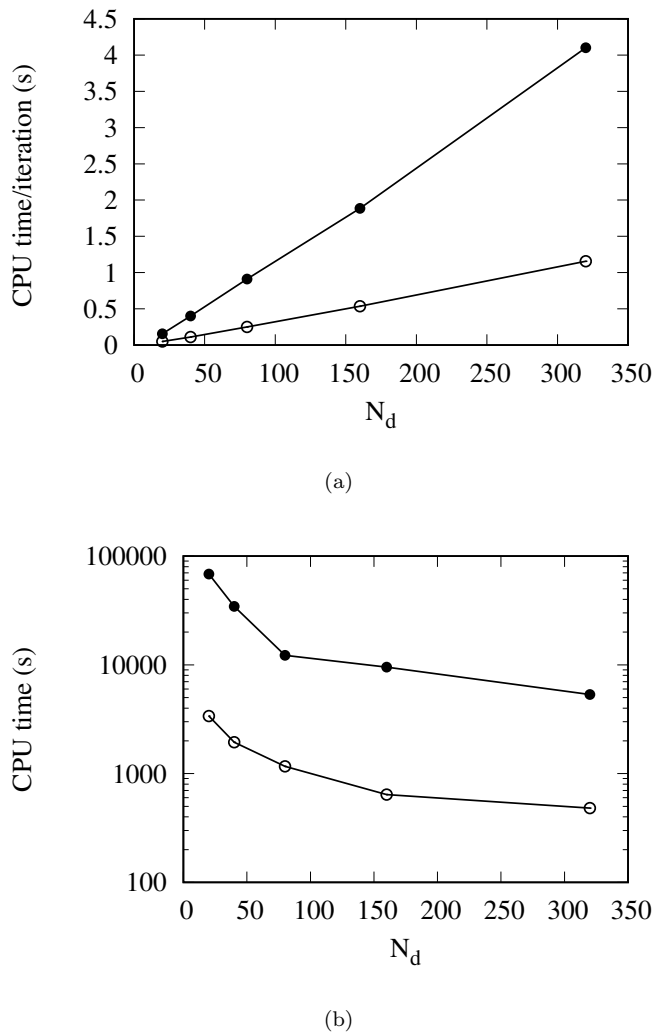


FIG. 1. a) Demonstration of the near linear scaling of the CPU time for the perturbative calculation per batch as a function of the number of determinants,  $N_d$ , sampled in each batch. The open and the filled circle are for the C<sub>2</sub> and F<sub>2</sub> molecules respectively with cc-pVQZ basis sets. b) CPU time in seconds required to reach a standard deviation of less than 0.1 mHa for various values of  $N_d$ . Note the initial rapid decrease in CPU time followed by a more gradual decrease at larger  $N_d$  values.

#### IV. IMPLEMENTATION

Here we briefly describe the implementation and the leading order cost of the various steps of the algorithm. In the variational stage there are three main operations, identifying the significant determinants to be included in the variational space, building the Hamiltonian matrix and diagonalizing the matrix. The cost of identifying important determinants is  $O(kN_v \ln(N_v) + kN_v \ln(N_p))$ , where  $N_v$  is the number of determinants in the variational space,  $k$  is the average number of  $H_{ai}$  elements for determinants  $D_i$  in  $\mathcal{V}$  that satisfy Eq. (1), and  $N_p = kN_v$  is the total number of new determinants that satisfy the

criterion in Eq. (1). The two terms in the cost function result from generating  $kN_v$  determinants and then doing a binary search of the list of the  $N_v$  existing variational determinants and the  $N_p$  newly generated determinants before including the determinant just generated in the newly generated determinant list.

In the current implementation we store all the nonzero elements of the Hamiltonian in memory using a *list of lists* (LIL) sparse storage format. In LIL format for each row we store a list containing the column index and the value of the nonzero Hamiltonian matrix elements.

The determinant labels are bit-packed strings that represent the occupancies of the up-spin ( $\alpha$ ) and the down-spin ( $\beta$ ) orbitals. To build the Hamiltonian efficiently, we first generate a list of all unique  $\beta$  strings and associated with each  $\beta$  string we store a list of all determinants in  $\mathcal{V}$  that have that  $\beta$  string. We also generate a list of all unique  $\alpha$  strings with  $N_\alpha - 1$  electrons and associated with each  $\alpha$  string we store a list of all determinants in  $\mathcal{V}$  that give the  $\alpha$  string on removing one  $\alpha$  electron. Here,  $N_\alpha$  is the number of  $\alpha$  electrons in our system. Determinants that are related to each other by double or single  $\alpha$  excitations have the same  $\beta$  string, and all the pairs of determinants that are related to each other with the remaining possible single or double excitations have the same  $\alpha$  string with  $N_\alpha - 1$  electrons. Hence to find the connected determinants, only the determinants in these two lists need to be considered rather than the entire set of  $N_v$  determinants in  $\mathcal{V}$ . Once the Hamiltonian is generated the Davidson algorithm is used to diagonalize it and the most expensive step there is the Hamiltonian wavefunction multiplication which costs  $O(kN_v)$ . Despite the fact that the Hamiltonian is sparse, building it is the most expensive part of the variational step, and storing it is currently the biggest memory bottleneck in the code. In the future we intend to implement the *direct* method for carrying out Hamiltonian wavefunction multiplication which does not require storing the Hamiltonian and can take less computer time as well<sup>43</sup>.

The stochastic perturbation step has two major components: sampling  $N_d$  determinants from the list of  $N_v$  variational determinants, and, identifying the determinants in  $\mathcal{C}$  that are connected to these  $N_d$  determinants and computing their contribution to the perturbative correction, Eq. (10). The  $N_d$  determinants are sampled using the Alias method<sup>44,45</sup>, which has an initial one-time memory cost of  $O(N_v)$ , and a subsequent cost of  $O(N_d)$  each time a sample is drawn. This method was used by some of us<sup>41</sup> for efficiently sampling determinants in the S-FCIQMC method. The CPU time for identifying the connected determinants along with their contributions is  $O(n^2 v^2 N_d \log(n^2 v^2 N_d) + n^2 v^2 N_d \log(N_v))$ , while the memory required is  $O(n^2 v^2 N_d + N_v)$ . Since the minimum required value of  $N_d$  is just two, the memory requirement for the stochastic perturbation theory is smaller than that of other parts of the calculation.

We have parallelized the entire code using hybrid OMP/MPI (open multiprocessing/message passing inter-

face) programming to make full use of the symmetric multiprocessor (SMP) architecture of most modern computers. A separate MPI process is initiated on each CPU and then each process forks into several threads (one for each computational core) on the CPU. The variational wavefunction is replicated on each CPU but a single copy is shared among the different threads on a CPU. As mentioned previously, the most memory intensive data structure is the LIL used to store the sparse Hamiltonian matrix. The list of nonzero matrix elements for each row of the Hamiltonian is distributed in round-robin fashion between the different CPUs. With this strategy the storage of the Hamiltonian and the computation of the Hamiltonian wavefunction multiplication is distributed approximately evenly between the different CPUs and threads. The perturbative step is embarrassingly parallel and no special strategy is needed to parallelize this step.

## V. BENCHMARKS

We perform frozen core calculations on a series of first row dimers including  $C_2$ ,  $N_2$ ,  $O_2$ ,  $NO$  and  $F_2$  with cc-pVDZ, cc-pVTZ and cc-pVQZ basis sets. Although the active spaces used for the first row diatomics have large Hilbert spaces, they are not a very stringent test for our theory because these molecules in their equilibrium geometry are not strongly correlated and traditional methods like CCSD(T) are cheap and reliable for such molecules. Thus, we also perform frozen-core calculations on the  $Cr_2$  dimer using cc-pVDZ, cc-pVTZ and cc-pVQZ bases, which have active spaces containing (12e, 68o), (12e, 118o) and (12e, 190o) respectively.  $Cr_2$  is well known for being very strongly correlated and most multi-reference methods can use no more than just the minimal active space and many of them fail to get even qualitatively correct dissociation curves. Finally, we also perform calculations on the Mn-Salen model complex which is a prototypical strongly correlated inorganic molecule containing open shell  $d$ -orbitals giving rise to nearly degenerate singlet and triplet ground states. For all the systems we obtain energies that are accurate to 1 mHa for the chosen basis; for the first-row dimers we compare to S-FCIQMC energies<sup>46</sup>, for the  $Cr_2$  dimer we perform internal convergence tests, and for Mn-Salen we compare to DMRG energies<sup>47</sup>.

### A. First row diatomics

In the variational calculations we start with a value of  $\epsilon_1$  during the first few iterations that is larger than its final value because the values of  $c_i$  tend to be larger in the initial iterations when there are few determinants in  $\mathcal{V}$ . For example, for the cc-pVQZ basis set, we successively reduce the value of  $\epsilon_1$  from  $10^{-3}$  to  $5 \times 10^{-4}$  to  $3 \times 10^{-4}$  and  $2 \times 10^{-4}$  Ha, and perform 3 iterations at each value. The cost of performing the first iteration at a value of

$\epsilon_1$  is larger than that for subsequent iterations because relatively few new determinants are introduced after the first iteration.

Table I shows benchmark calculations on the first row dimers using a single node containing two Intel® Xeon® E5-2680 v2 processors of 2.80 GHz each and 128 gigabyte memory. Among these calculations  $F_2$  had the largest active space containing 14 electrons in 108 orbitals (14e, 108o) with a Hilbert space containing over  $10^{20}$  determinants. On a single node it required slightly more than 4 minutes to get the energy converged to better than 1 mHa. It is not possible to perform the calculations for the larger systems and basis sets on a single node with the original algorithm because the cost of storing all the determinants in the space of connections  $\mathcal{C}$  that contribute to the perturbative corrections is prohibitive. Interestingly, with our implementation the cost of obtaining sub-mHa accuracy in energy using the stochastic method for even the smallest system considered here,  $C_2$  with DZ basis, is less than that for the original deterministic algorithm.

As expected, these calculations can be done even more efficiently, if the Hartree-Fock orbitals are replaced by natural orbitals from some approximate correlated theory. For example the last three rows of Table I show that the calculations on the  $F_2$  dimer, for all three basis sets, can be run with a larger  $\epsilon_1$  resulting in more than a factor of 2 speed up when MP2 natural orbitals are used.

### B. $Cr_2$ dimer

The  $Cr_2$  dimer is well known to be a very challenging system for most electronic structure methods. We perform frozen core calculations by including all the available virtual orbitals in the active space with a bond length of 1.68 Å using the cc-pVDZ-DK, cc-pvTZ-DK and cc-pVQZ-DK basis sets. The relativistic effects are included using the second order Douglas-Kroll-Hess Hamiltonian. All calculations are performed using natural orbitals obtained by first performing a short unconverged FCIQMC calculation. The active spaces with the DZ, TZ and QZ basis sets contained (12e, 68o), (12e, 118o) and (12e, 190o) respectively, with the largest Hilbert space containing more than  $10^{21}$  determinants. Table II shows that although the variational energies are far from convergence, the total energies including the perturbative corrections converge rapidly as  $\epsilon_1$  is reduced. Another point to note is that the variational wavefunction contains determinants with all excitation orders, all the way up to the maximum possible of 12. Hence a CI expansion that is truncated at some excitation level is not adequate.

Since the stochastic error of the perturbative calculation decreases as  $1/\sqrt{N_b}$ , where  $N_b$  is the number of batches used in the perturbative calculations, the time for the perturbative calculation can be greatly reduced if a larger statistical error is acceptable. For instance, in the

TABLE I. Ground state s-HCI energies of the  $C_2$ ,  $N_2$ ,  $O_2$ ,  $NO$  and  $F_2$  molecules with bond lengths of 1.2425, 1.0977, 1.2075, 1.1508 and 1.4119 Å respectively and DZ, TZ and QZ basis sets. The variational cutoff,  $\epsilon_1$ , the number of determinants in the variational space,  $N_v$ , the variational energy, Var, and the total energy obtained after the stochastic perturbative correction is added, PT, are shown. All perturbative calculations were performed with an  $\epsilon_2 = 10^{-8}$  Ha and  $N_d = 200$ . The last three rows show that the s-HCI calculations converge with much looser  $\epsilon_1$  threshold when MP2 natural orbitals, rather than canonical HF orbitals, are used. The final three columns show the wall time in seconds required to perform the variational and the perturbative parts of the calculation. All the results obtained from perturbation theory (PT) agree within error bars with the results published previously using FCIQMC<sup>46</sup>. Each calculation was performed on a single node (see text for details).

Molecule	Basis	Sym	$\epsilon_1(Ha)$	$N_v$	Energy (Ha)		Walltime (sec)		
					Var	PT	Var	PT	Total
$C_2$	DZ	$^1A_{1g}$	$5 \times 10^{-4}$	28566	-75.7217	-75.7286(2)	1	2	3
$C_2$	TZ	$^1A_{1g}$	$3 \times 10^{-4}$	142467	-75.7738	-75.7846(3)	10	6	16
$C_2$	QZ	$^1A_{1g}$	$2 \times 10^{-4}$	403071	-75.7894	-75.8018(4)	63	11	75
$N_2$	DZ	$^1A_{1g}$	$5 \times 10^{-4}$	37593	-109.2692	-109.2769(1)	1	3	4
$N_2$	TZ	$^1A_{1g}$	$3 \times 10^{-4}$	189080	-109.3608	-109.3748(6)	14	8	22
$N_2$	QZ	$^1A_{1g}$	$2 \times 10^{-4}$	499644	-109.3884	-109.4055(9)	77	18	95
$O_2$	DZ	$^1A_{1g}$	$5 \times 10^{-4}$	52907	-149.9793	-149.9878(2)	2	3	4
$O_2$	TZ	$^1A_{1g}$	$3 \times 10^{-4}$	290980	-150.1130	-150.1307(8)	24	7	30
$O_2$	QZ	$^1A_{1g}$	$2 \times 10^{-4}$	770069	-150.1541	-150.1748(9)	131	30	161
$NO$	DZ	$^1B_1$	$5 \times 10^{-4}$	48305	-129.5881	-129.5997(3)	2	3	5
$NO$	TZ	$^1B_1$	$3 \times 10^{-4}$	227004	-129.6973	-129.7181(9)	30	12	42
$NO$	QZ	$^1B_1$	$2 \times 10^{-4}$	606381	-129.7311	-129.7548(9)	207	60	267
$F_2$	DZ	$^1A_{1g}$	$5 \times 10^{-4}$	68994	-199.0913	-199.1001(7)	2	3	5
$F_2$	TZ	$^1A_{1g}$	$3 \times 10^{-4}$	395744	-199.2782	-199.2984(9)	37	8	46
$F_2$	QZ	$^1A_{1g}$	$2 \times 10^{-4}$	1053491	-199.3349	-199.3590(9)	216	41	257
<i>Natural Orbitals</i>									
$F_2$	DZ	$^1A_{1g}$	$1 \times 10^{-3}$	16824	-199.0871	-199.0994(4)	0	3	3
$F_2$	TZ	$^1A_{1g}$	$5 \times 10^{-4}$	141433	-199.2787	-199.2972(7)	11	16	27
$F_2$	QZ	$^1A_{1g}$	$5 \times 10^{-4}$	221160	-199.3355	-199.3590(9)	34	79	113

$Cr_2$  calculation with QZ basis set with  $\epsilon_1 = 0.5 \times 10^{-4}$  Ha, if a stochastic error of 1 mHa is acceptable, then the perturbative step would take 8763 seconds, which is comparable to the variational calculation that took 6267 seconds. It is noteworthy that the cost of the perturbative correction relative to the variational calculations decreases as the size of the variational space  $\mathcal{V}$  increases. In fact, the CPU time required to reach a fixed statistical error is virtually insensitive to the size of the variational space for a given basis set. The time for achieving an uncertainty of 1 mHa is shown in the second last column of Table II. It changes very little with the size of  $\mathcal{V}$ . This indicates that much larger calculations could be performed if the memory and CPU cost of the variational step were reduced, potentially by using the direct CI method<sup>43</sup>.

### C. Mn-Salen

Finally, we demonstrate that s-HCI can be used to calculate the active space energy of a prototypical strongly correlated molecule like Mn-Salen ( $MnClO_3N_2C_8H_{10}$ ) (see Figure 2) very quickly. Mn-Salen derivatives such as Jacobson's catalyst are used to catalyze enantioselective epoxidation of olefins. Despite their widespread use and importance, the mechanism of the catalysis reaction is not known and has spawned a series of theoretical studies<sup>48–58</sup>. Recently, some of us performed DMRG-SCF calculations<sup>47</sup> on the model cluster with the cc-pVDZ basis set using an active space of (28e, 22o). The initial orbitals were obtained by using the HOMO-13 to LUMO+7 canonical Hartree Fock orbitals, which were subsequently optimized using the DMRG-SCF method. Here we perform the s-HCI calculations on the converged orbitals obtained at the end of the converged DMRG-

TABLE II. Ground state s-HCI energies of the  $\text{Cr}_2$  molecule with bond length 1.68 Å with DZ, TZ and QZ basis sets. The various columns in the table have the same meaning as those in Table I with the exception of the two additional columns. The column labeled PT(1) under Walltime shows the time it would have taken to perform the stochastic perturbative step if we terminated the calculation after obtaining an uncertainty of 1 mHa. The final column labeled #Nodes shows the number of nodes that were used to perform these calculations.

Basis	Sym	$\epsilon_1(\text{Ha})$	$N_v$	Energy (Ha)		Walltime (sec)				#Nodes
				Var	PT	Var	PT	Total	PT(1)	
DZ	$^1A_{1g}$	$2.5 \times 10^{-4}$	602984	-2099.4518	-2099.4859(6)	66	6458	6524	2172	1
DZ	$^1A_{1g}$	$1.0 \times 10^{-4}$	2261194	-2099.4665	-2099.4869(7)	355	5279	5634	2891	1
DZ	$^1A_{1g}$	$0.8 \times 10^{-4}$	3117630	-2099.4693	-2099.4873(3)	685	28115	28799	2702	1
TZ	$^1A_{1g}$	$0.8 \times 10^{-4}$	6268840	-2099.5051	-2099.5280(6)	1245	7736	8981	2879	4
TZ	$^1A_{1g}$	$0.5 \times 10^{-4}$	12756099	-2099.5113	-2099.5292(6)	3197	8559	11756	3081	4
TZ	$^1A_{1g}$	$0.4 \times 10^{-4}$	17798876	-2099.5166	-2099.5295(5)	5717	14099	19816	3525	4
QZ	$^1A_{1g}$	$0.8 \times 10^{-4}$	9516339	-2099.5246	-2099.5578(4)	2164	54764	56928	8762	8
QZ	$^1A_{1g}$	$0.5 \times 10^{-4}$	19500559	-2099.5315	-2099.5562(7)	6267	16999	23265	8763	8

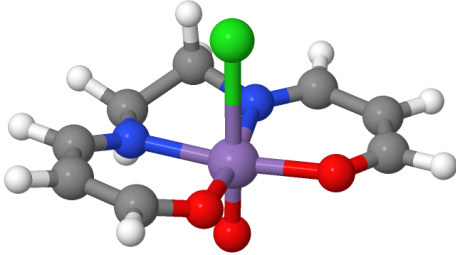


FIG. 2. The geometry of the chlorine containing neutral oxo-Mn(salen) cluster was optimized with (10e,10o)-CASSCF and a 6-31G\* basis by Ivanic et al.<sup>48</sup>. The Mn, Cl, N, O, C, H atoms are shown in purple, green, blue, red, grey and white respectively.

TABLE III. Comparison of the DMRG and s-HCI energies (E+2251) Ha of the singlet and triplet states of Mn-Salen. DMRG was performed with an  $M = 2000$  and s-HCI was performed with  $\epsilon_1 = 2 \times 10^{-4}$  Ha,  $\epsilon_2 = 1 \times 10^{-8}$  Ha and  $N_d = 200$ . The wall time needed to perform the s-HCI calculation on a single compute node is shown in the final column(see text for additional details).

Sym	$N_v$	Energy (Ha)			Walltime (s)
		Var	PT	DMRG	
				Ref. 47	
$^1A$	232484	-0.7880	-0.7980(7)	-0.7991	65
$^3A$	208334	-0.7910	-0.7994(8)	-0.8001	54

SCF calculations. The results in Table III show that both the singlet and the triplet energies converge to better than 1 mHa accuracy in only 65 seconds on a single node.

## VI. CONCLUSIONS

We have introduced a stochastic implementation of multireference Epstein-Nesbet perturbation theory, for computing the expectation value of the perturbative correction to the variational energy of a multi-determinant wavefunction without storing all the contributing determinants. In addition to completely removing the memory bottleneck, the stochastic algorithm is faster than the fully deterministic algorithm even for relatively small systems, if a stochastic noise of 1 mHa is acceptable.

Our method is capable of efficiently computing the correlation energies of very large active spaces, as we have demonstrated by computing the energies of the challenging, multireference systems Mn-Salen (28e, 22o) and  $\text{Cr}_2$  (12e, 190o). For all systems studied we obtained correlation energies accurate to within 1 mHa. In the case of the first-row dimers and Mn-Salen we compared to FCIQMC and DMRG energies in the literature. For  $\text{Cr}_2$  there are no published values, but one of the positive features of our method is that one can reliably check the convergence within the method itself.

Having removed the memory bottleneck in the perturbative step, the largest memory requirement comes from storing the Hamiltonian in the variational space. The next step is to create an efficient method for obtaining the variational wavefunction without storing the Hamiltonian. Other research directions include the optimization of the orbitals within the CAS space, and the calculation of excited states.

## ACKNOWLEDGMENTS

The calculations made use of the facilities of the Max Planck Society's Rechenzentrum Garching. SS acknowledges the startup package from the University of Col-



orado. AAH and CJU were supported in part by NSF grant ACI-1534965.

- 
- \* sanshar@gmail.com  
 † aah95@cornell.edu  
 ‡ a.alavi@fkf.mpg.de  
 § cyrusumrigar@gmail.com
- <sup>1</sup> J. Olsen, B. O. Roos, P. Jorgensen and H. J. A. Jensen, *J. Chem. Phys.* **89**, 2185 (1988).
  - <sup>2</sup> P. A. Malmqvist, A. Rendell and B. O. Roos, *J. Phys. Chem.* **94**, 5477 (1990).
  - <sup>3</sup> D. Ma, G. Li Manni and L. Gagliardi, *J. Chem. Phys.* **135**, 044128 (2011).
  - <sup>4</sup> S. R. White, *Phys. Rev. Lett.* **69**, 2863 (1992).
  - <sup>5</sup> S. R. White, *Phys. Rev. B* **48**, 10345 (1993).
  - <sup>6</sup> G. K.-L. Chan and S. Sharma, *Annu. Rev. Phys. Chem.* **62**, 465 (2011).
  - <sup>7</sup> G. H. Booth, A. J. Thom and A. Alavi, *J. Chem. Phys.* **131**, 054106 (2009).
  - <sup>8</sup> D. Cleland, G. H. Booth and A. Alavi, *J. Chem. Phys.* **132**, 041103 (2010).
  - <sup>9</sup> F. R. Petruzielo, A. A. Holmes, H. J. Changlani, M. P. Nightingale and C. J. Umrigar, *Phys. Rev. Lett.* **109**, 230201 (2012).
  - <sup>10</sup> H.-J. Werner and E. A. Reinsch, *J. Chem. Phys.* **76**, 3144 (1982).
  - <sup>11</sup> P. J. Knowles and H.-J. Werner, *Theoretica chimica acta* **84**, 95 (1992).
  - <sup>12</sup> K. R. Shamasundar, G. Knizia and H.-J. Werner, *J. Chem. Phys.* **135**, 054101 (2011).
  - <sup>13</sup> K. Andersson, P. A. Malmqvist, B. O. Roos, A. J. Sadlej and K. Wolinski, *J. Phys. Chem.* **94**, 5483 (1990).
  - <sup>14</sup> R. F. Fink, *Chemical Physics* **356**, 39 (2009).
  - <sup>15</sup> C. Angeli, R. Cimiraglia, S. Evangelisti, T. Leininger and J.-P. Malrieu, *J. Chem. Phys.* **114**, 10252 (2001).
  - <sup>16</sup> K. Hirao, *Chemical Physics Letters* **190**, 374 (1992).
  - <sup>17</sup> D. I. Lyakh, M. Musiał, V. F. Lotrich and R. J. Bartlett, *Chemical reviews* **112**, 182 (2012).
  - <sup>18</sup> F. A. Evangelista, M. Hanauer, A. Köhn and J. Gauss, *J. Chem. Phys.* **136**, 204108 (2012).
  - <sup>19</sup> S. Sharma, G. Jeanmairet and A. Alavi, *J. Chem. Phys.* **144**, 034103 (2016).
  - <sup>20</sup> S. Sharma and A. Alavi, *J. Chem. Phys.* **143**, 102815 (2015).
  - <sup>21</sup> B. Huron, J. Malrieu and P. Rancurel, *J. Chem. Phys.* **58**, 5745 (1973).
  - <sup>22</sup> R. J. Buenker and S. D. Peyerimhoff, *Theor. Chim. Acta* **35**, 33 (1974).
  - <sup>23</sup> S. Evangelisti, J.-P. Daudey and J.-P. Malrieu, *Chemical Physics* **75**, 91 (1983).
  - <sup>24</sup> R. J. Harrison, *J. Chem. Phys.* **94**, 5021 (1991).
  - <sup>25</sup> M. M. Steiner, W. Wenzel, K. G. Wilson and J. W. Wilkins, *Chem. Phys. Lett.* **231**, 263 (1994).
  - <sup>26</sup> W. Wenzel, M. Steiner and K. G. Wilson, *Int. J. Quantum Chem.* **60**, 1325 (1996).
  - <sup>27</sup> F. Neese, *J. Chem. Phys.* **119**, 9428 (2003).
  - <sup>28</sup> M. L. Abrams and C. D. Sherrill, *Chem. Phys. Lett.* **412**, 121 (2005).
  - <sup>29</sup> L. Bytautas and K. Ruedenberg, *Chem. Phys.* **356**, 64 (2009).
  - <sup>30</sup> F. A. Evangelista, *J. Chem. Phys.* **140**, 054109 (2014).
  - <sup>31</sup> P. J. Knowles, *Mol. Phys.* **113**, 1655 (2015).
  - <sup>32</sup> J. B. Schriber and F. A. Evangelista, *J. Chem. Phys.* **144**, 161106 (2016).
  - <sup>33</sup> N. M. Tubman, J. Lee, T. Y. Takeshita, M. Head-Gordon and K. B. Whaley, *J. Chem. Phys.* **145**, 044112 (2016).
  - <sup>34</sup> W. Liu and M. R. Hoffmann, *J. Chem. Theory Comput.* **12**, 1169 (2016).
  - <sup>35</sup> M. Caffarel, T. Applecourt, E. Giner and A. Scemama, Using CIPSI nodes in diffusion Monte Carlo, <http://arxiv.org/abs/physics.chem-ph/1607.06742v2>.
  - <sup>36</sup> The  $N_v$  determinants in the variational space are connected to  $\mathcal{O}(n^2 v^2 N_v)$  determinants in the perturbative space with non-zero Hamiltonian matrix elements, where  $n$  is the number of electrons and  $v$  is the number of virtual orbitals. For a relatively conservative number of  $n = 12$ ,  $v = 50$  and  $N_v = 10^7$ , the perturbative space will contain over  $10^{12}$  determinants, requiring over 10 terabyte memory. The original HCI algorithm reduces this storage requirement by orders of magnitude by only requiring the storage of determinants  $D_a$  for which  $|H_{ai}c_i| > \epsilon_2$  for at least one determinant  $D_i \in \mathcal{V}$ . Nevertheless, for many systems this is the most memory intensive part of the original algorithm.
  - <sup>37</sup> A. A. Holmes, N. M. Tubman and C. J. Umrigar, *J. Chem. Theory Comput.* **12**, 3674 (2016).
  - <sup>38</sup> P. S. Epstein, *Phys. Rev.* **28**, 6956 (1926).
  - <sup>39</sup> R. K. Nesbet, *Proc. R. Soc. London, Ser. A* **230**, 312 (1955).
  - <sup>40</sup> A. Schäfer, H. Horn and R. Ahlrichs, *J. Chem. Phys.* **97**, 2571 (1992).
  - <sup>41</sup> A. A. Holmes, H. J. Changlani and C. J. Umrigar, *J. Chem. Theory Comput.* (2016).
  - <sup>42</sup> The logarithmic corrections come from having to do binary searches to check whether a determinant is already present in the variational space, or, in the space of connected determinants that have already been generated.
  - <sup>43</sup> P. Knowles and N. Handy, *Chemical Physics Letters* **111**, 315 (1984).
  - <sup>44</sup> A. J. Walker, *ACM Trans. on Math. Software (TOMS)* **3**, 253 (1977).
  - <sup>45</sup> R. A. Kronmal and A. V. Peterson Jr, *Amer. Statist.* **33**, 214 (1979).
  - <sup>46</sup> D. Cleland, G. H. Booth, C. Overy and A. Alavi, *J. Chem. Theory Comput.* **8**, 4138 (2012).
  - <sup>47</sup> S. Sharma, G. Knizia, S. Guo and A. Alavi (2016).
  - <sup>48</sup> J. Ivanic, J. R. Collins and S. K. Burt, *J. Phys. Chem. A* **108**, 2314 (2004).
  - <sup>49</sup> F. Teixeira, R. A. Mosquera, A. Melo, C. Freire and M. N. D. S. Cordeiro, *Phys. Chem. Chem. Phys.* **16**, 25364 (2014).
  - <sup>50</sup> C. J. Stein and M. Reiher, *J. Chem. Theory Comput.* (2016).
  - <sup>51</sup> T. Bogaerts, S. Wouters, P. VanDerVoort and V. VanSpeybroeck, *ChemCatChem* **7**, 2711 (2015).
  - <sup>52</sup> D. Ma, G. Li Manni and L. Gagliardi, *J. Chem. Phys.* **135**, 044128 (2011).
  - <sup>53</sup> J. S. Sears and C. D. Sherrill, *J. Chem. Phys.* **124**, 144314 (2006).

- <sup>54</sup> S. Wouters, T. Bogaerts, P. Van Der Voort, V. Van Speybroeck and D. Van Neck, *J. Chem. Phys.* **140**, 241103 (2014).
- <sup>55</sup> C. Linde, B. Åkermark, P.-O. Norrby and M. Svensson, *Journal of the American Chemical Society* **121**, 5083 (1999).
- <sup>56</sup> Y. G. Abashkin, J. R. Collins and S. K. Burt, *Inorganic Chemistry* **40**, 4040 (2001).
- <sup>57</sup> Y. G. Abashkin and S. K. Burt, *J. Phys. Chem. B* **108**, 2708 (2004).
- <sup>58</sup> Y. G. Abashkin, J. R. Collins and S. K. Burt, *Inorganic Chemistry* **40**, 4040 (2001).